Jurnal Pendidikan Matematika
Volume 17, No. 2, May 2023 pp. 247-264
P-ISSN: 1978-0044, E-ISSN: 2549-1040, DOI: https://doi.org/10.22342/jpm.17.2.2023.247-264
Website: https://ejournal.unsri.ac.id/index.php/jpm
Accredited by SINTA 2: http://sinta2.ristekdikti.go.id/journals/detail?id=1811

Developing Statistical Literacy Through Tasks: An Analysis of Secondary School Mathematics Textbooks

Ezra Putranda Setiawan[1], Heru Sukoco[2], Anggit Reviana Dewi Agustyani[3]

[1, 2, 3]Department of Mathematics Education, Universitas Negeri Yogyakarta, Jl. Colombo 1, Depok, Sleman, Indonesia
Email: ezra.ps@uny.ac.id

Abstract

Statistical literacy is an essential competence that people must face in the era of big data and Society 5.0. In Indonesia and several countries, statistics is taught as a list of competencies in mathematics subject in primary and secondary schools. This study aimed to identify whether statistical tasks in higher secondary school mathematics textbooks support statistical literacy. The qualitative data were collected via deductive content analysis, with a specific framework, toward five Compulsory Mathematics textbooks used in Indonesia. We found that statistical exercises and problems in these textbooks were dominated by data analysis type, emphasizing calculating statistics from raw data and its modification. Regarding data visualization, almost all textbooks introduced the histogram and ogive, while some also introduced the boxplot, stem-and-leaf plot, and dot plot. Improvement could be made by adding more exercises and problems related to the interpretation of statistics, evaluation of statistical results, and comparison of statistics from several data groups.

Keywords: Statistical Literacy, Task, Textbooks, Visualization

Abstrak

Literasi statistika merupakan salah satu kemampuan yang harus dimiliki setiap orang pada era *Big Data* dan masyarakat 5.0. Di Indonesia dan beberapa negara lainnya, statistika dipelajari sebagai sejumlah kompetensi pada mata pelajaran Matematika di jenjang pendidikan dasar dan menengah. Penelitian ini bertujuan mengetahui sejauh mana aktivitas-aktivitas pada buku matematika jenjang Sekoah Menengah Atas (SMA) mendukung kompetensi literasi Statistika ini. Data kualitatif dikumpulkan melalui analisis isi secara deduktif (dengan kerangka tertentu) terhadap lima buku pelajaran Matematika wajib yang dipergunakan di Indonesia. Diperoleh hasil bahwa soal-soal latihan dan permasalahan statistika pada buku-buku ini didominasi oleh analisis data, dengan penekanan pada perhitungan statistik dari data mentah dan modifikasinya. Terkait visualisasi data, hampir semua buku memuat penjelasan mengenai histogram dan ogive, sedangkan beberapa buku di antaranya juga memberikan penjelasan terkait diagram kotak-garis, diagram batang-daun, dan diagram titik. Perbaikan lebih lanjut dapat dilakukan dengan menambah latihan-latihan terkait interpretasi statistik, evaluasi hasil statistik, serta pembandingan nilai statistik pada data dari beberapa kelompok.

Kata kunci: Buku teks, Literasi Statistika, Soal, Visualisasi.

How to Cite: Setiawan, E. P., Sukoco, H., & Agustyani, A. R. D. (2023). Developing statistical literacy through tasks: An analysis of secondary school mathematics textbooks. *Jurnal Pendidikan Matematika*, *17*(2), 247-264.

INTRODUCTION

The rapid growth of computing technology, the Internet of Things (IoT), artificial intelligence, and big data brings the fourth industrial revolution to societies worldwide. Along with this growth, various kinds of data have been generated at a high rate (Yaqoob et al., 2016). Big data, which has become a popular term, promote data usage in many situations to make suitable decisions. As the data can be found everywhere, the need for statistical literacy is increased. Citizens should be able to utilize and understand statistics to filter, critique, analyze, and interpret the available data to make decisions (Jones & Jacobbe, 2014). They also need to understand statistics published in public situations such as

Received January 18, 2023; Revised May 25, 2023; Accepted May 30, 2023

mass media and official statistics (Engel, 2017; Porciúncula et al., 2018).

What is statistical literacy? Statistics is known as the art of learning from data (Ross, 2010), while literacy means the ability to read and write (Montoya, 2018). The term statistical literacy was suggested by Walker (1951) as the ability to communicate statistical information. Following Shield & Dole, (2013), statistical literacy can be seen through various perspectives such as chance-based, correlation-based, and misusage-based. Statistical literacy is also known as 'transnumerative thinking,' where people will be capable of making sense of and using a different representation of data to make sense of the situation among them (Chick et al., 2005). Garfield & DelMas (2010) stated that statistical literacy involves understanding and using statistics' primary language and tools.

Due to its importance, statistical literacy has become part of the mathematical curriculum in various countries (Garfield & DelMas, 2010), including Brazil (Campos et al., 2008) and the USA (Weiland, 2017). A guideline for developing a curriculum with statistical literacy competence has been suggested by (Schield, 2004). In Indonesia, statistical literacy has been studied in mathematics from the primary school level up to the senior secondary school level (Setiawan, 2019).

Even though students have been learning statistics for many years, several studies show that undergraduate students still face problems with it. In Indonesia, Jatisunda et al., 2020) presented that students from the undergraduate Public Administration program in a private university in West Java still have an unsatisfactory level of statistical literacy. A similar result was also found in students from the undergraduate mathematics education program in Banten, Indonesia (Khaerunnisa & Pamungkas, 2017). Setiawan & Sukoco (2021) find that some undergraduate students in Indonesia still face difficulties in determining the descriptive statistics and data visualization that should be used, especially when two groups should be compared. In other countries, studies by Sotos et al. (2007), Shaughnessy (2007), and Vanhoof et al. (2006) show that many students are statistically illiterate and ill-prepared for tertiary-level statistics even after completing one or several courses in statistics.

What causes learning problems in statistical literacy? Research suggests that students' problem with statistical literacy relates to how they study statistics in school. It is known that most of the statistics training in school focuses on the procedural and computational aspects of statistics rather than on conceptual understanding (Hardin et al., 2015; Shaughnessy & Pfannkuch, 2002). The traditional formula-based approach to teaching statistics, which emphasizes accumulating statistical knowledge, memorizing facts and formulas, and following rules and standard procedures, did not yield sufficient competence in statistical literacy (Batanero et al., 2011). Moreover, in the era of Artificial Intelligence, many procedures such as data collection, statistical calculation, and data visualization can be carried out automatically in real-time situations (Peres et al., 2018; Roh et al., 2021; Toasa et al., 2018). The remaining task for the people become to understand the data and take any decision based on the data. In these two areas, statistical literacy had an essential role.

Is students' problem in statistical literacy related to the textbooks used in their schools? It is widely understood that books play an essential role in mathematics teaching since it is used by teachers to plan the lesson (Alajmi, 2012) and determine the teaching strategies (Reys et al., 2004). Variability among textbooks and how it is used might yield different opportunities for learning (Shield & Dole, 2013; Thompson et al., 2012). Tasks or activities in any mathematics textbooks also contribute to the learning opportunities for students to develop understanding (Yang et al., 2017).

Several studies on the statistical content of secondary school mathematics textbooks have been carried out in Indonesia, for example, by Setiawan (2019, 2020) and Purnama et al. (2020). However, it is still unknown whether some activities or tasks in these books are sufficient to build statistical literacy competence. Here, statistical literacy should not be seen only as an ability to answer questions by calculating some descriptive statistics but as an ability to conduct statistical problem-solving procedures. Following the framework of Jones et al. (2015), statistical problem-solving is a process that consists of formulating questions, collecting data, analyzing data, and interpreting the results. An in-depth discussion about statistical problem-solving is available in (Marriott et al., 2009).

This study analyzed several Indonesian mathematics textbooks at the senior secondary school level. We expected to identify whether these textbooks' tasks, exercises, and problems are sufficient to support students' statistical literacy through statistical problem-solving.

METHODS

This study was focused on how statistical tasks were provided and arranged in mathematics textbooks for senior secondary school. Therefore, we used a qualitative content analysis method known as textbook analysis. Following Charalambous et al. (2010), this study employed horizontal and vertical analysis of textbooks. The horizontal analysis was used to provide a general overview, while the vertical one was carried out to obtain in-depth and comprehensive information about the statistical content of these books.

Textbooks Selection

From Setiawan (2019), the statistics-related competencies were presented in the twelfth grade or the final grade of senior secondary school. In this study, we focus on the compulsory mathematics program, which means that this competence is studied by students from all streams, namely the mathematics and natural science stream, social science stream, and language stream. A discussion about statistical competencies for students from mathematics and natural science streams can be seen in Setiawan (2020).

The subjects of this study were several mathematical textbooks written for twelfth-grade students (i.e. the highest grade in senior secondary school) based on the 2018 revisions of the 2013 curriculum. All textbooks presented in Table 1 were the newest edition of mathematical textbooks for the twelfth grade still used in senior secondary schools in Indonesia. As the contents of the Indonesian curriculum undergo a simplification due to the Covid-19 pandemic, there is no revision on mathematics textbooks for this grade. On the other hand, books for the 12th-grade students based on the recent 'Kurikulum Merdeka' is not yet available.

Table 1. List of 12th-grade mathematical textbooks and their code

Title	Publisher	Published Year	Code
Matematika XII	Ministry of Education and Culture	2018	A
Matematika 3A	Erlangga (private publisher)	2018	B
Matematika 3	Erlangga (private publisher)	2018	C
Jelajah Matematika 3	Yudhistira (private publisher)	2018	D
Perspektif Matematika 3	Tiga Serangkai (private publisher)	2019	E

Frameworks

The analysis of tasks in mathematics textbooks can be done in several ways, for example, by the cognitive level framework (e.g., Giani et al., 2015) or cognitive domain and dimension (e.g., Cahyono & Adilah, 2016). Regarding statistical contents, a framework based on a statistical problem-solving procedure was introduced by Jones & Jacobbe (2014). This framework is proposed under the Guidelines for Assessment and Instruction in Statistics Education (GAISE). In this framework, there are four types of statistical tasks, namely (1) formulate questions, (2) collect data, (3) analyze data, and (4) interpret results and predictions.

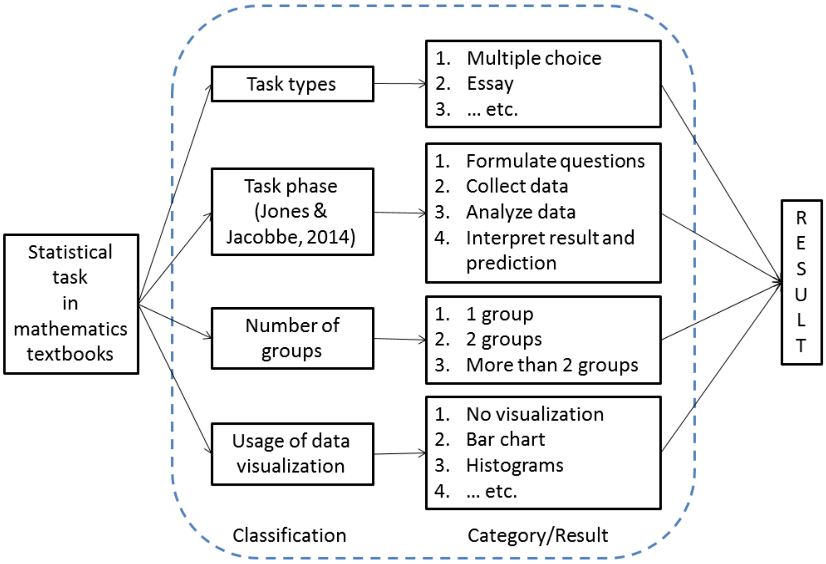


Figure 1. Classification framework of statistical tasks used in this study

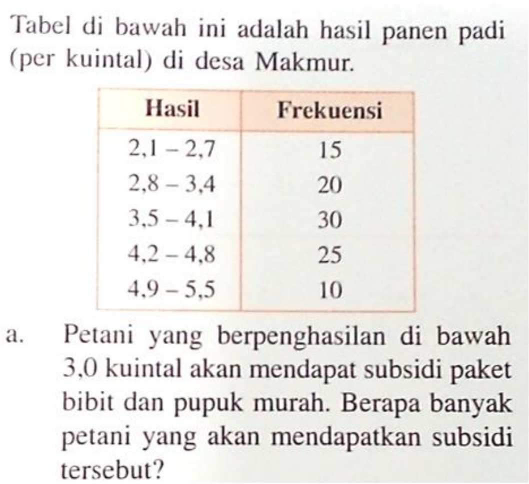
In this study, we used this framework for two reasons. First, this framework provided clear guidelines (Jones & Jacobbe, 2014). The guideline also contained several examples of problems from each type of task, so we only needed to compare them to the problems that appeared in Indonesian mathematics textbooks. Second, this framework was also used by Jones et al. (2015), Yanık et al. (2017), as well Jones & Basyal (2019) for different levels of mathematics textbooks from various areas.

As a consequence, the obtained result can be compared easily.

Furthermore, we believe that making a comparison between groups and visualizing the data is an essential skill for students (Setiawan & Sukoco, 2021). Therefore, we propose two additional criteria for analyzing statistical tasks: the number of variables and groups and the use of data visualization. Consequently, the classification framework used in this study is rather complicated, as presented in Figure 1.

Data Collection

Following the study by Jones et al. (2015), the unit of study here is statistical tasks, which can be any questions, problems, or exercises in the textbook. We define a task as the smallest marked divisions, as presented in Figure 2.


Tabel di bawah ini adalah hasil panen padi (per kuintal) di desa Makmur.

Hasil	Frekuensi
2,1 – 2,7	15
2,8 – 3,4	20
3,5 – 4,1	30
4,2 – 4,8	25
4,9 – 5,5	10

a. Petani yang berpenghasilan di bawah 3,0 kuintal akan mendapat subsidi paket bibit dan pupuk murah. Berapa banyak petani yang akan mendapatkan subsidi tersebut?

[C, page 115]
Table below presents the weight of rice produced in Makmur village. Suppose that farmers with result less than 3 will get subsidized packages of seeds and cheap fertilizer. How many farmers that will obtain the packages?

Figure 2. Sample of statistical task found in the textbook. Although it is related to one table or dataset, each number of questions is considered a separate statistical task

The data collection procedure was carried out as follows. Finding the chapter(s) related to statistics in these textbooks, we made a list containing all statistical tasks. Following the framework presented in Figure 1, we examined each statistics task in detail to classify them according to four criteria, namely (1) the type of task, (2) the phase of the statistical task based on the framework, (3) the number of groups, and (4) the use of data visualization. Since the first, third, and fourth criteria were objective, the classification of each criterion was done by each researcher. However, we realized that the second criterion was somewhat subjective rather than objective. Therefore, the second criteria were examined for each task by each researcher separately. The different classification was then discussed to obtain a consensus.

In descriptive-qualitative research, we presented the result using proper data display and provided some discussion about the results.

RESULTS AND DISCUSSION

In this section, we present and discuss the result of this study. To maintain the path, we present the result in four parts, namely (1) an overview of studied textbooks, (2) an overview of statistical tasks, (3) an analysis of the phase of the task, and (4) data visualization and groups. The last subsection is devoted to discussion and implications.

Overview of Studied Textbooks

According to the decree of the Indonesian Minister of Education and Culture number 24/2016 about the core and basic competence in Mathematics for Senior Secondary School, there are two competencies related to statistics for students in 12th grade as follows.

> (3.2) Determine and analyze central tendency and variability measure of data which are presented in the frequency distribution table and histogram,
> (4.2) Solve problems related to displaying data from measurement and counting in a frequency distribution table and histogram.

Apart from the statistical-related competence, the other competencies studied in this grade are related to geometry (solid geometry), counting rules, and probability theory. Solid geometry is discussed in Chapter 1 before statistics, whereas the counting rules and probability theory are the last chapters following the statistics chapter.

All investigated textbooks' content related to these competencies is arranged in one chapter, as presented in Table 2. However, there are differences in the number of subchapters. Book A only separates the content into two subchapters: data display and centrality and dispersion measure of grouped data. On the other hand, books D and E separate the content into five sub-chapters by separating the central tendency measure of grouped data and the dispersion measure. These books also add a section on data position, which discusses the data's quartile and percentile.

Table 2. The statistics chapter title, number of sub-chapter, and number of pages

Book	Title of statistics chapter	# of sub chapter	# of pages	Proportion
A	Statistics	2	55	21.5%
B	Centrality and dispersion measure of grouped data	3	90	39.5%
C	Statistics	5	104	38.4%
D	Statistics	5	51	42.5%
E	Descriptive Statistics	5	81	35.8%

According to Table 2, the proportion of statistical content in textbook A seems lower than in the other textbook because it provides an additional chapter named ‘kesebangunan’ (congruence) as enrichment material. When this chapter was not counted, the proportion of statistical content among the whole book A became 31.2%. Therefore, statistical contents comprise at least 31% of the mathematical textbooks for Grade XII senior secondary school.

Overview of Statistical Tasks

The statistical tasks in all mathematics textbooks studied here can be separated into two groups, namely within-chapter tasks and end-of-chapter tasks, as presented in Table 3. Within-chapter tasks are usually given at the end of a sub-chapter and can be seen as a formative assessment.

Table 3. Position and types of statistical tasks in the textbooks

Book	Within-chapter tasks			End-of-chapter tasks		
	Multiple choice	Essay	Other	Multiple choice	Essay	Other
A	0	29	1	0	15	0
B	175	214	18	40	0	0
C	120	215	8	40	55	3
D	0	78	0	20	16	0
E	0	110	1	25	11	0

Table 3 shows that most of the statistical tasks in these textbooks are in the form of open questions (essays), especially within-chapter tasks. There are more within-chapter tasks than end-of-chapter ones. Among these textbooks, textbook A provides the smallest number of exercises, whereas B and C account for many exercises. In addition to the multiple-choice and essay questions, there are other exercise types found in these textbooks, for example, short answers (in A, B, and C) and true/false questions (in E).

In addition to the exercises, Textbooks C and E also provide statistical projects. This project guides the students to collect the data on their weight and height, then visualize them and calculate several descriptive statistics using a scientific calculator or a computer. However, both books did not explain the steps for creating such visualization using the computer. Interestingly, Textbook C also offers another project in which students should get secondary data from a small company and present them as histograms, frequency distribution tables, and graphics.

Analysis of Task Phase

All statistical tasks in these books were separated into four phases of statistical problem-solving, based on the framework proposed by Jones & Jacobbe (2014). The results are presented in Table 4 as follows.

Table 4. Distribution of tasks addressing phases of statistical problem-solving

Book	Total # of statistical tasks	Number of tasks in phase			
		Formulate questions	Collect data	Analyse data	Interpret result and prediction
A	45	3	1	34	14
B	447	0	0	439	8
C	441	0	3	428	29
D	114	0	2	111	0
E	147	0	0	142	0

*a statistical task can be categorized into more than one phase.

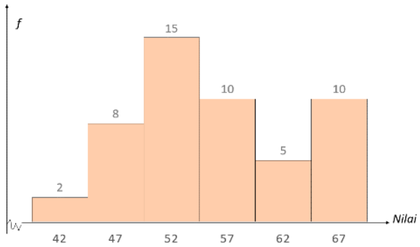
As presented in Table 4, more than 70 percent of statistical exercise in these textbooks is in data analysis. Several statistical tasks ask the student to formulate questions or collect data. Statistical task on formulating questions is only found in Textbook A, which is related to the histogram, polygon frequency, and ogive of a dataset. Data collection tasks were presented in two textbooks, consisting of instructions for collecting data (in A) and theoretical explanations about methods for collecting data (in D). The interpretation and prediction task is only found in Textbooks A, B, and C, with a small proportion, compared to the data analysis task. In other words, most of the tasks did not relate to the ability to make a conclusion or prediction.

Since most of the statistical tasks are on analyzing data, we used the approach by (Jones et al., (2015) by classifying the tasks into more specific criteria. Jones et al. (2000) propose four categories, namely (1) read a display, (2) perform a mathematical calculation, (3) construct a display, and (4) use other statistical reasoning. However, we modified these criteria due to many tasks in our samples. First, we separated the category 'perform a mathematical calculation' into four different categories, namely

1. calculation statistics from raw data or tables,
2. determination of missing data from statistics,
3. symbolic calculation of statistics and properties of statistics, and
4. modification of statistics calculation.

The last category, '*modification of statistics calculation,'* was used for any statistics calculation whose aim was not to find the statistics (e.g., mean) or the missing data. Next, since 'calculation of statistics from a chart' overlapped with 'read a display, they were grouped into one criterion. Last, we separated '*constructs a display'* into two categories, namely (1) creating a table and (2) visualizing data. The classification frameworks then look as in Table 5.

Table 5. Types of statistical tasks in the phase of analyzing data

<table>
<tr><th>Category</th><th>Examples</th></tr>
<tr><td>1. Create a table from raw data or diagrams.</td><td>[D, page 29] Write a relative frequency distribution table and a cumulative relative table from the following data.
21 35 32 21 22
40 22 31 34 38</td></tr>
<tr><td>2. Calculating a statistics from raw data or table.</td><td>[B, page 67] Look at the frequency distribution table below.<table><tr><td>Value</td><td>51-60</td><td>61-70</td><td>71-80</td><td>81-90</td><td>91-100</td></tr><tr><td>Frequency</td><td>5</td><td>7</td><td>14</td><td>8</td><td>6</td></tr></table>The mean of the above data is ….</td></tr>
<tr><td>3. Calculating a statistics from a chart.</td><td>[B, page 74] The mean value of data presented by the histogram below is ….
</td></tr>
<tr><td>4. Modification of statistics calculation.</td><td>[E, page 56] A company give its workers' salary once in two weeks. The mean of salary is IDR 250,000.00. The mean of salary for male workers is IDR 260,000.00, while the mean salary for female workers is IDR 210,000.00. Find the ratio between male and female workers in the company.</td></tr>
<tr><td>5. Symbolic calculation of statistics and properties of statistics.</td><td>[E, page 57] Given data from a measurement, namely x_1, x_2, x_3, …, x_{10} with mean x_0. If these data are changed into $\frac{x_1}{3}+1$, $\frac{x_2}{3}+3$, $\frac{x_3}{3}+5$, …, $\frac{x_{10}}{3}+19$, find the mean of the changed data!</td></tr>
<tr><td>6. Determine missing data.</td><td>[B, page 62] The mean value of 7, 5, 13, *x* and 9 is 10. The value of *x* is ….</td></tr>
<tr><td>7. Visualizing data.</td><td>[C, page 70] The profit of a marketing company during 2007-2016 is presented in table below (all number is in million rupiah).<table><tr><td>Year</td><td>2007</td><td>2008</td><td>2009</td><td>2010</td><td>2011</td><td>2012</td><td>2013</td><td>2014</td><td>2015</td><td>2016</td></tr><tr><td>Profit</td><td>2</td><td>3</td><td>3,5</td><td>5</td><td>7</td><td>8</td><td>7,5</td><td>9</td><td>10</td><td>8</td></tr></table>Present the above data in forms of a line chart.</td></tr>
<tr><td>8. Miscellaneous problem.</td><td>(All of statistical task related with data analysis that could not be classified into the other categories).</td></tr>
</table>

Based on the criteria presented in Table 5, the types of data analysis tasks in these textbooks are presented in Figure 3. It can be inferred that the calculation of statistics from raw data dominates the data analysis problem in all textbooks, followed by calculating statistics from charts/diagrams and then visualizing data. Textbooks B, D, and E provide more statistical tasks on modification of statistics calculation, which are not available in the other textbooks. Otherwise, Textbook C contains the highest number of statistical tasks on interpreting charts, with some unique problems that are not found in the other textbooks.

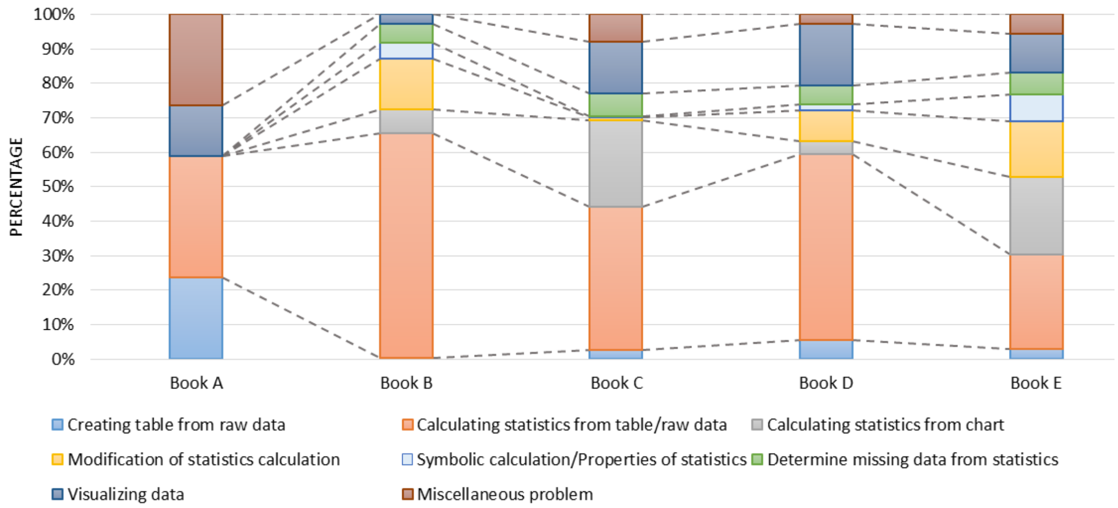


Figure 3. Distribution of data analysis tasks in mathematics textbooks based on the criteria on Table 5, shown as a percentage

Regarding the type of statistical data analysis tasks, Figure 3 shows that Textbook A only provides four types which are minimum among others. Besides, Textbooks D and E provide eight types, while B and C provide six and seven types, respectively. About half of the statistical tasks in Textbooks B and D belong to one category: calculating statistics from the table or raw data. Compared to the others, Textbooks B and C have a higher number of tasks, so a similar percentage should not be interpreted as a similar number of statistical tasks.

The presence of miscellanea problems can be explained as follows. Since the number of data analysis tasks in Textbook A is deficient, the percentage of miscellanea problems seems higher than the others. These problems consist of evaluating the frequency distribution table (page 54) and the usage of a statistical distribution (page 77). In Textbook D, the miscellanea related to steps on creating grouped distribution (page 28) and guessing four numbers are based on three statistics (page 58), while in Textbook E, this category contains problems regarding missing data with algebraic calculation (page 56) and creating a distribution frequency table from diagrams (pages 100, 101, and 113).

Analysis of Data Visualization and Groups

The above classification implies that there are two types of tasks related to data visualization: the calculation of statistics from a chart and the construction of a chart. As seen in Figure 4, Textbook C contains the highest number of data visualization with various types, while Textbook A only provides a small number of data visualization.

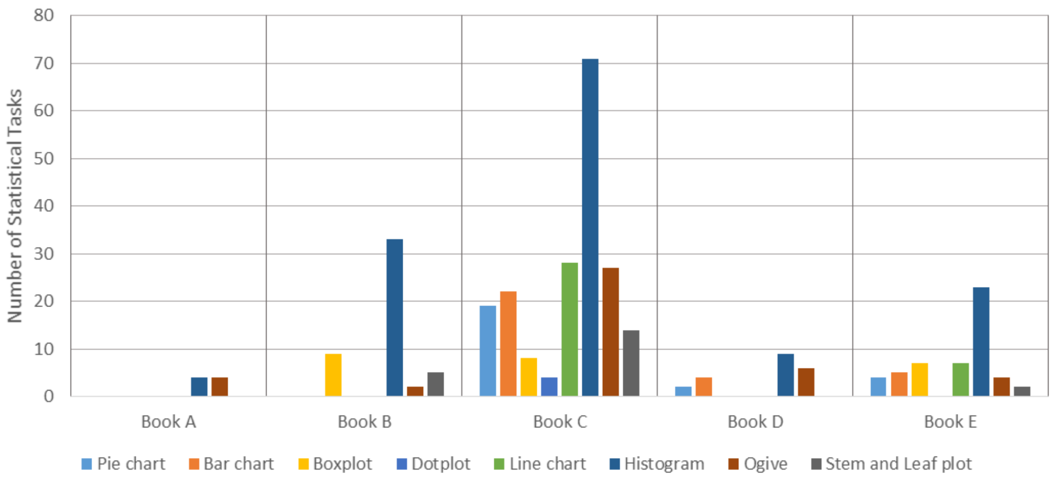
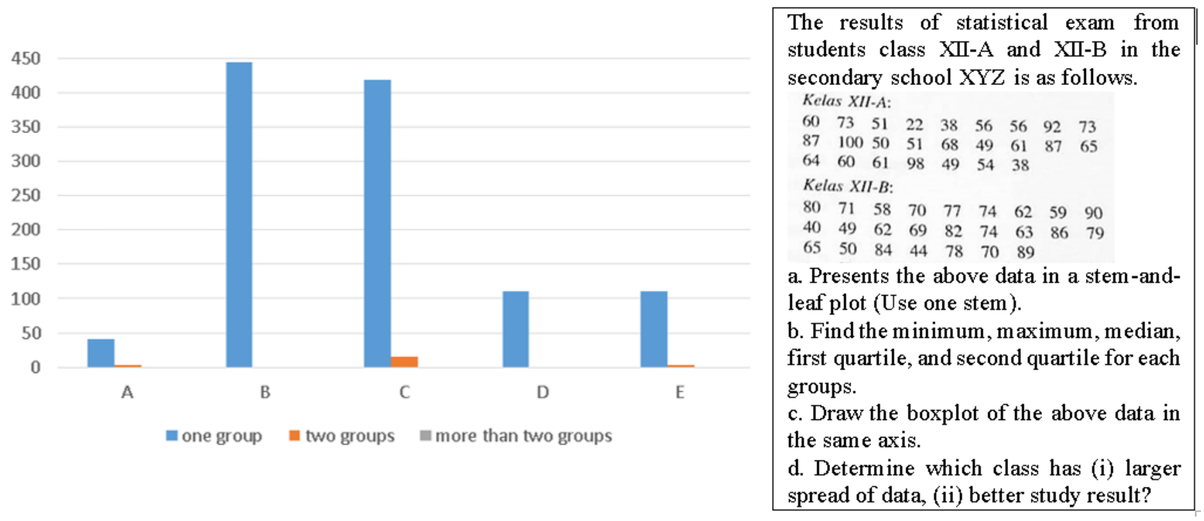


Figure 4. Usage of data visualization on statistical tasks

Figure 4 lets us examine the types of data visualization used in these textbooks. The histogram is the most frequent type of chart presented in statistical tasks, sometimes equipped with polygon frequency. Ogive, which can be constructed based on a cumulative frequency table, is found in four books. Some books also provide other types of data visualization such as bar charts (C, D, and E), stem and leaf plots (B, C, and E), boxplots (B and E), line charts (C and E), and then dot plot (C only). Regarding the boxplot (or box-and-whisker plot), it should be noted that these textbooks do not mention the inner fence or outer fence or classify an observation as an outlier. The example of a boxplot presented in these textbooks does not have any observation outside the whisker.



The results of statistical exam from students class XII-A and XII-B in the secondary school XYZ is as follows.

Kelas XII-A:

60	73	51	22	38	56	56	92	73
87	100	50	51	68	49	61	87	65
64	60	61	98	49	54	38		

Kelas XII-B:

80	71	58	70	77	74	62	59	90
40	49	62	69	82	74	63	86	79
65	50	84	44	78	70	89		

a. Presents the above data in a stem-and-leaf plot (Use one stem).
b. Find the minimum, maximum, median, first quartile, and second quartile for each groups.
c. Draw the boxplot of the above data in the same axis.
d. Determine which class has (i) larger spread of data, (ii) better study result?

Figure 5. The number of groups in statistical tasks on each textbook (left), sample statistical task found in Textbook C that involves more than one group (right).

The number of groups in each task is related to students' opportunity to make a comparison between the groups and among the members of the same group. We check whether the statistical tasks account for one or more groups (e.g., the height of male vs. female students) and present the result in Figure 5.

From Figure 5, we can see that most of the statistical tasks in these textbooks are dominated by the use of one group of data. Moreover, two textbooks only give tasks with one group of data, and none of them contains a comparison between more than two groups. As presented in the right part of Figure 5, problems related to data for more than one group would help the students understand how to compare the data correctly. However, these sampled textbooks rarely contain this kind of statistical task.

Across the 12th-grade mathematical textbooks used in this study, the statistical contents are found on more than 30% of the pages. This result seems high, as mathematics textbooks at a lower level in other countries might contain less than 20% (Jones et al., 2015; Jones & Jacobbe, 2014; Pickle, 2012). The large portion of statistics might relate to the arrangement of contents in the curriculum, which only contains four competencies for this grade. In Indonesia, statistical competence in the 12th grade can be seen as a continuation of statistical content in the 9th grade (Setiawan, 2019).

This study finds that the 'data analysis' tasks dominate the statistical tasks in almost all sampled textbooks. Furthermore, at least 20% of data analysis tasks is about calculating statistics from raw data and table. A similar situation can be found in the U.S. mathematics textbooks for elementary students (Jones et al., 2015), Nepali mathematics textbooks for Grades 4-10 (Jones & Basyal, 2019), as well as US High School Textbooks (Tran & Tarr, 2018). In Indonesia, this dominance might be caused by the curriculum, which only mentions central tendency, variability measure, and data display, which means a procedure for data analysis. It should be realized that the steps of 'formulating questions,' 'collecting data' and 'interpreting results and prediction' are not explicitly stated in the curriculum. Surprisingly, we still find two textbooks that provide statistical projects that include collecting data, although the students are not given any opportunity to determine the data they want to collect.

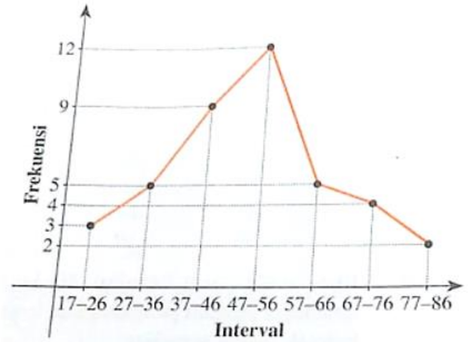
Regarding the data analysis task, complicated calculations are not always related to complex statistical procedures. Some statistical tasks might be easy to answer but need some understanding of the actual situation or context of the data. In contrast, some tasks need tricky (perhaps challenging) calculations but only contain numbers without context. The balance between these types of questions should be made carefully.

Our study finds that there are several questionable statistical tasks in these textbooks. First, there are statistical tasks that do not give any understanding to the student except to calculate any statistics from a list of numbers (Figure 6a). Since any descriptive statistics can be calculated using any software, the presence of many statistical tasks of this type is ineffective. Another questionable task is calculating statistics from data already visualized in the forms of ogive or polygon frequency (Figure 6b). Although the calculation might be done correctly, such a procedure is rarely found in actual statistical activities. In our daily lives, someone who needs to use secondary data would try to get the numerical forms of the data instead of calculating the statistics from the visualization. Such tasks might confuse the students and make the course seems more difficult.

(a)
Find the range, interquartile range (IQR), quartile deviation, average deviation, and standard deviation from the following data.

a. 2, 4, 7, 6, 8, 8, 3, 2.
b. 1, 2, 3, 4, 5, 4, 3, 2, 3.
c. 7, 7, 8, 8, 8, 9, 9, 10, 11, 12, 13, 6.
d. 12, 14, 15, 15, 17, 17, 13, 12, 9, 11, 15, 18.
e. 40, 41, 35, 29, 27, 26, 24, 17, 15.

(b)
Look at the following polygon frequency.



Calculate the first quartile (Q1), median (Q2), and third quartile (Q3) from the above data.

Figure 6. Sample statistical data analysis tasks found in one translated textbook

Regarding the data visualization, only histograms (with/without polygon frequency) and ogive are presented in all the sampled textbooks. Again, these results might be related to the competence in the curriculum documents, which mention displaying the data from measurement and counting in a frequency distribution table and histogram. Since the frequency distribution table can be used to calculate descriptive statistics, a boxplot can be created. Boxplot will help make a comparison between two or more groups of data. When the ungrouped data are used, a stem-and-leaf plot also can be used to perform this comparison. An emphasis should be given that histogram, ogive, boxplot, and stem-and-leaf plot are suitable for numerical data, so students are familiar with them (Setiawan & Sukoco, 2021). On the other hand, the number of statistical tasks related to the pie chart, bar charts, and line charts can be reduced since these competencies have been studied at the junior secondary school level (Setiawan, 2019).

The proper use of diagrams and descriptive statistics becomes crucial when we need to compare data from two or more groups as well as two or more variables, such as the test score in algebra and geometry or the test score of male and female students. However, this study found that more than 75% of statistical tasks in the mathematical textbook for higher secondary schools only contain univariate (one variable) data. Besides, we only found a small number of statistical tasks guiding the students to compare two or more data groups and interpret the result. This situation is in line with Setiawan & Sukoco's (2021) findings that some students who graduated from secondary school still had difficulties determining accurate statistics or proper data visualization when they are asked to compare two groups of data.

Based on the above explanations, several recommendations can be made as follows. First, mathematics textbooks should contain more statistical tasks in comparing data groups and using more than one variable. Statistical tasks may contain more 'real datasets' with more data. In this situation, the calculation of descriptive statistics and production of diagrams can be done using the spreadsheet (i.e., Excel and OpenOffice Calc), and statistical software (i.e., R, Python, Minitab, and the like). The competence in the Indonesian mathematics curriculum for secondary school does not restrict the number of variables, groups, or data used in the analysis of central tendency or variability measurement. Consequently, these improvements can be executed on the statistical task without violating the curriculum.

This study still has some limitations, as follows. First, it should be realized that textbooks were not the only source for studying statistics. Consequently, further study can be done to find how secondary school mathematics teachers taught the statistical content to the students, for example, whether they taught how to interpret and draw boxplots or use any computer software. Second, this study did not check whether the teacher was able to provide better problems or exercises that could improve students' statistical literacy. Although they used specific textbooks, it was likely that the teacher could develop new problems or exercise by themselves. Last, because the textbooks might not reflect students' statistical literacy, further study should be carried out on the students especially to measure their statistical literacy using a valid and reliable instrument.

CONCLUSIONS

We analyze statistical tasks on five mathematics textbooks for Indonesia's 12th senior secondary schools. Most of the statistical tasks in these textbooks address data analysis competence, focusing on calculating statistics from raw data and diagrams. Regarding data visualization, most of them present histograms and ogive, while only several textbooks introduce more types of diagrams such as box plots and stem-and-leaf plots. Only a few statistical tasks are related to the comparison between groups of data. Moreover, there are a few numbers of statistical tasks that ask the student to make questions, collect the data, and make interpretations or predictions based on data, as these competencies are not stated explicitly in the curriculum.

To help the students develop their statistical literacy, senior secondary school mathematics textbook writers should give more statistical tasks, various data visualization types, and some questions of more than one data group.

ACKNOWLEDGMENTS

The authors thank the faculty librarians for providing textbooks, UNY Language Center for proof-reading services, and to anonymous reviewers for suggestions to improve this manuscript.

REFERENCES

Alajmi, A. H. (2012). How do elementary textbooks address fractions? A review of mathematics textbooks in the USA, Japan, and Kuwait. *Educational Studies in Mathematics*, *79*(2), 239–261. https://doi.org/10.1007/s10649-011-9342-1

Batanero, C., Burrill, G., & Reading, C. E. (2011). *Teaching statistics in school mathematics - Challenges for teaching and teacher education. A joint ICMI/IASE study: The 18th ICMI study* (1st ed.). Springer. https://doi.org/10.1007/978-94-007-1131-0

Cahyono, B., & Adilah, N. (2016). Analisis soal dalam buku siswa matematika kurikulum 2013 kelas VIII semester I berdasarkan dimensi kognitif dari TIMSS. *Jurnal Review Pembelajaran Matematika*, *1*(1), 86–98. https://doi.org/10.15642/jrpm.2016.1.1.86-98

Campos, T. M. M., da Silva, C. B., & Cazorla, I. M. (2008). Statistical literacy in Brazil in high and middle school: An analysis of official documents. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. https://iase-web.org/documents/papers/rt2008/T1P6_Campos.pdf?1402524989

Charalambous, C., Delaney, S., Hsu, H.-Y., & Mesa, V. (2010). A comparative analysis of the addition and subtraction of fractions in textbooks from three countries. *Mathematical Thinking and Learning*, *12*, 117–151. https://doi.org/10.1080/10986060903460070

Chick, H. L., Pfannkuch, M., & Watson, J. M. (2005). Transnumerative thinking: Finding and telling stories within data. *Curriculum Matters*, *1*, 87–108. https://doi.org/10.18296/cm.0063

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16*(1), 44–49. https://doi.org/10.52041/serj.v16i1.213

Garfield, J., & DelMas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics*, *32*(1), 2–7. https://doi.org/https://doi.org/10.1111/j.1467-9639.2009.00373.x

Giani, Zulkardi, & Hiltrimartin, C. (2015). Analisis tingkat kognitif soal-soal buku teks matematika kelas VII berdasarkan taksonomi Bloom. *Jurnal Pendidikan Matematika*, *9*(2), 78–98. https://doi.org/10.22342/jpm.9.2.2125.78 - 98

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., Baumer, B., Hall-Holt, O., Murrell, P., Peng, R., Roback, P., Temple Lang, D., & Ward, M. D. (2015). Data science in statistics curricula: Preparing students to "think with data." *The American Statistician*, *69*(4), 343–353. https://doi.org/10.1080/00031305.2015.1077729

Jatisunda, M. G., Nahdi, D. S., & Suciawati, V. (2020). Kemampuan Literasi Statistika Mahasiswa Adminitrasi Publik. *SJME (Supremum Journal of Mathematics Education)*, *4*(2), 134–146. https://doi.org/10.35706/sjme.v4i2.3488

Jones, D. L., & Basyal, D. (2019). An analysis of the statistics content in Nepali school textbooks. *Mathematics Education Forum Chitwan*, *4*(4), 21–34. https://doi.org/10.3126/mefc.v4i4.26356

Jones, D. L., Brown, M., Dunkle, A., Hixon, L., Yoder, N., & Silbernick, Z. (2015). The statistical content of elementary school mathematics textbooks. *Journal of Statistics Education*, *23*(3), 1–22. https://doi.org/10.1080/10691898.2015.11889748

Jones, D. L., & Jacobbe, T. (2014). An analysis of the statistical content in textbooks for prospective elementary teachers. *Journal of Statistics Education*, *22*(3), 1–12. https://doi.org/10.1080/10691898.2014.11889713

Jones, G. A., Thornton, C., Langrall, C., Mooney, E., Perry, B., & Putt, I. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning*, *2*(4), 269–307. https://doi.org/10.1207/S15327833MTL0204_3

Khaerunnisa, E., & Pamungkas, A. S. (2017). Profil kemampuan literasi statistis mahasiswa jurusan pendidikan matematika Universitas Sultan Ageng Tirtayasa. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *6*(2), 246–255. https://doi.org/10.24127/ajpm.v6i2.970

Marriott, J., Davies, N., & Gibson, L. (2009). Teaching, learning and assessing statistical problem solving. *Journal of Statistics Education*, *17*(1), 1–18. https://doi.org/10.1080/10691898.2009.11889503

Montoya, S. (2018). Defining literacy. *GAML Fifth Meeting*. https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/12/4.6.1_07_4.6-defining-literacy.pdf

Peres, R. S., Dionisio Rocha, A., Leitao, P., & Barata, J. (2018). IDARTS – Towards intelligent data analysis and real-time supervision for industry 4.0. *Computers in Industry*, *101*, 138–146. https://doi.org/10.1016/j.compind.2018.07.004

Pickle, M. C. C. (2012). *Statistical content in middle grades mathematics textbooks*. https://digitalcommons.usf.edu/etd

Porciúncula, M., Samá, S., de Arimatéa Rocha, C., & de Carvalho, J. I. F. (2018). Every citizen needs to know statistics! What are we doing? Brazilian research in statistics education. In A. J. Ribeiro, L. Healy, R. E. de S. R. Borba, & S. H. A. A. Fernandes (Eds.), *Mathematics Education in Brazil : Panorama of Current Research* (pp. 249–263). Springer International Publishing. https://doi.org/10.1007/978-3-319-93455-6_13

Purnama, A., Wijaya, T., Dewi, S., & Zulfah, Z. (2020). Analisis buku siswa matematika SMA dari Indonesia dan China pada materi peluang dan statistik. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, *4*, 813–822. https://doi.org/10.31004/cendekia.v4i2.305

Reys, B. J., Reys, R. E., & Chavez, O. (2004). Why mathematics textbooks matter. *Educational Leadership*, *61*(5), 61–66. https://www.researchgate.net/publication/234571079

Roh, Y., Heo, G., & Whang, S. E. (2021). A survey on data collection for machine learning: A big data - AI integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, *33*(4), 1328–1347. https://doi.org/10.1109/TKDE.2019.2946162

Ross, S. M. (2010). Introductory Statistics. In S. M. Ross (Ed.), *Introductory Statistics (Third Edition)*. Academic Press. https://doi.org/10.1016/B978-0-12-374388-6.00019-3

Schield, M. (2004). *Statistical Literacy Curriculum Design*. www.StatLit.org/pdf/2004SchieldIASE.pdf

Setiawan, E. P. (2019). Analisis muatan literasi statistika dalam buku teks matematika Kurikulum 2013. *Pythagoras: Jurnal Pendidikan Matematika*, *14*(2), 163–177. https://doi.org/10.21831/pg.v14i2.28558

Setiawan, E. P. (2020). Introducing statistical inference to senior high school students: a textbook analysis. *Journal of Physics: Conference Series*, *1663*(1), 012014. https://doi.org/10.1088/1742-6596/1663/1/012014

Setiawan, E. P., & Sukoco, H. (2021). Exploring first year university students' statistical literacy: A case on describing and visualizing data. *Journal on Mathematics Education*, *12*(3), 427–448. https://doi.org/10.22342/JME.12.3.13202.427-448

Shaughnessy, J. M. (2007). Research on statistical learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Information Age Publishing.

Shaughnessy, J. M., & Pfannkuch, M. (2002). How faithful is Old Faithful? Statistical thinking: A story of variation and prediction. *The Mathematics Teacher*, *95*(4), 252–259. https://doi.org/10.5951/MT.95.4.0252

Shield, M., & Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics*, *82*(2), 183–199. https://doi.org/10.1007/s10649-012-9415-9

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98–113. https://doi.org/10.1016/j.edurev.2007.04.001

Thompson, D., Senk, S., & Johnson, G. (2012). Opportunities to learn reasoning and proof in high school mathematics textbooks. *Journal for Research in Mathematics Education*, *43*, 253–295. https://doi.org/10.5951/jresematheduc.43.3.0253

Toasa, R., Maximiano, M., Reis, C., & Guevara, D. (2018). Data visualization techniques for real-time information — A custom and dynamic dashboard for analyzing surveys' results. *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–7. https://doi.org/10.23919/CISTI.2018.8398641

Tran, D., & Tarr, J. E. (2018). Examination of bivariate data tasks in US high school textbooks through the statistical investigation and cognitive demands frameworks. *International Journal of Science and Mathematics Education*, *16*(8), 1581–1603. https://doi.org/10.1007/s10763-017-9851-1

Vanhoof, S., Castro Sotos, A. E., Onghena, P., Verschaffel, L., Van Dooren, W., & Van den Noortgate, W. (2006). Attitudes toward statistics and their relationship with short-and long-term exam results. *Journal of Statistics Education*, *14*(3), null-null. https://doi.org/10.1080/10691898.2006.11910588

Walker, H. M. (1951). Statistical literacy in the social sciences. *The American Statistician*, *5*(1), 6–12. https://doi.org/10.1080/00031305.1951.10481912

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, *96*(1), 33–47. https://doi.org/10.1007/s10649-017-9764-5

Yang, D. C., Tseng, Y. K., & Wang, T. L. (2017). A comparison of geometry problems in middle-grade mathematics textbooks from Taiwan, Singapore, Finland, and the United States. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(7), 2841–2857. https://doi.org/10.12973/eurasia.2017.00721a

Yanık, H. B., Özdemir, G., & Çevirgen, A. (2017). Ortaokul matematik ders kitaplarında yer alan veri işlemeye yönelik görevlerin incelenmesi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *18*(2), 45–61. https://doi.org/10.17679/inuefd.323407

Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, *36*(6, Part B), 1231–1247. https://doi.org/10.1016/j.ijinfomgt.2016.07.009